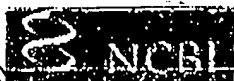


09/17/55



PlumMet

## FOUR

BLAST

CMM

## Taxonomy

## Structure

Match: 5 Mismatch: 2 gap open: 1 gap extension: 2  
x dropoff: 100 expect: 10.000 wordsize: 7 Filter ☒ Align:

Sequence 1 of 4538106 Sequence 3 from Patent WO9740167. Length 8209 (1 .. 8209)

Sequence 2 of 9630311 Gibbon leukemia virus, complete genome Length 8088 (1 .. 8088)

-----The statistics (hitscore and expect value) is calculated based on the size of nr database

----- If protein translation is reversed, please repeat the search with reverse strand of the query sequence

Score = 1583 bits (21250), Expect = 0.0  
Identities = 5776/8943 (64%), Gaps = 1625/8943 (18%)  
Strand = Plus / Plus

**Query:** 2 tgggtgtacgaactgtg-ggcccccag--cgcgct-tg--gaataaaaaatcctcttgctgctt  
| | | | | | | | | | | | | | | | | | | | | |  
**sp1ct:** 12 taqaq-a-qacctgagccgcgcgqtaccgc-tgtgtccaataaaa--cctcttgctgat

```
Query:      56  gcatcaaga-cgcct--tctcgtgagtgat--ttg--ggggtgcgcct--ct-tccga-
              ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:     67  gaatcaggagccg-tgctatcgt---tcttcctggggaggggttctcctaactat-tgac
```

```

Query:      104  -gccccggac-gaggggggattgt-tcttttactggcccttcatattgggtgcgttggg-cgagg
           ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:     122  cgccc--actcggggg--tctcacatttgggggc--tc---gtccc--ggatcga

```

```

Query:      160 aa----atcctgcgaccacc--cctracacc--c--gag--aacccgactcgg--ag-gta
             ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Sbjct:      160 aacccgaccacgggacccaccgacc--caccaacggggaggttaa--g-c-tggccagcg-a

```

```

Query:      205  aag--g-g-atc-ccctt-tg-g---aa--cnnnnnnnnnnnnnc-ggcc--g-gcgtc--
             | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:      220  ccgttgtgtgtctcgccttctgtgcctaaagtc-cgtaatctcgacgtccttgtgtgtctc

```

```

Query:      247  ---tctgt-tctga---gt--gtctg--t---tttcgggtga-tgcgcgc---tttcgg
              ||||| |||||  ||  |||||  ||  ||  |||||  |||||  |||||
Sbjct:      279  gcttctctgtctgaagaccgtaacctctgactgcaccc--gtaagtcgcgcgcatttttttg

```

```
Query:                287   ttctgcag-ctgtctctctcagaccgt-aagga--ctg-gag-gac-tgtga-t--c-agca
                        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Sbjct:                337   tt-cagctctct--tc-cga---gtcaatcactctgcgaagtgcagctgtgagtagcggaac
```

```

Query:      336  gacggtgctaggaggatcacaggct-gccaccctgggggacgccccggagggt---ggggg-
              |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Sbjct:      390  gacgtgttcgg-ggctcaccgcctgggtaatccaggagacgtcccagga--tcaggggag

```

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?1>

21/02/01

Query:		392	gagccagggaacgcctgggtggtcctactagtctggttcagaggaccagatttctgtttgaa
Sbjct:		447	ga-ccagggaacgcctgggtggac--c--c-tcggt-a-a----cg-ggtc-gttg-tgac
Query:		452	gcaaaagcttc-cccctccgcg---g-c-cg-tccgactct-ttgtccctgct-tgtggaa
Sbjct:		492	cogatttcatcgcccgcttg-gtaagacgcgctctgaatctgatt--ct-ctctctcg--
Query:		503	gacgcgggacgggtcg-cg-tgtgtctcg-atctgttgggttctgtctcgctg---tgtctt
Sbjct:		546	gtcgc---c---tcgcccgctctctgggttctctttttgtttc-gtttc-tggaaaacctc
Query:	VIRAL CORE PROTEIN	557	tgtcttgt-gc-gtccttcttac--agt-tttaatatgggac-agacagtgaactcccc 1 M G Q T V T T P
Sbjct:	gag polyprotein	598	tg---tgtcacagctcttctctccccaatcatcaatatgggacaagaataat-tctacccc 1 M G Q D N S T P
Query:	VIRAL CORE PROTEIN	611	ccttagt-t---tgactctcgaccattggactgaagtfagatccagggtcataatttgt 9 L S L T L D H W T E V R S R A H N L
Sbjct:	gag polyprotein	654	---ta-tctccctcactctaatactcagagagatgtgagaacaagggtcaccaatctat 9 I S L T L N H W R D V R T R A H N L
Query:	VIRAL CORE PROTEIN	667	cagttcagggttaagaagggaaccttggcagacrttctgtgcctctgaatggccaacattcg 27 S V Q V K K G P W Q T F C A S E W P T F
Sbjct:	gag polyprotein	710	ccgtggaaatcaaaaagggaataatggcagactttctgttctccgagtggtggccacattcg 27 S V E I K K G K W Q T F C S S E W P T F
Query:	VIRAL CORE PROTEIN	727	atgttgatggccatcagaggggaccccttat---tetgaattaccotsgotgttaaag 47 D V G W P S E G T F N S E I I L A V K
Sbjct:	gag polyprotein	770	gggtgggtgggggggggaactcttaattctctgtcatt-t---ttgcagttaa-- 47 G V G W P P E G T F N L S V I F A V K
Query:	VIRAL CORE PROTEIN	784	caatcatt---tttca-gactggaccggctctcactcctgatcaggag--ccctatatoc 66 A I I F O T G P G S H P D Q E P Y I
Sbjct:	gag polyprotein	825	--aagattgtcttccaggga---ga-acgggggacatccggacca--agttccatatatcg 66 K I V F Q E N G G H P D Q V P Y I
Query:	VIRAL CORE PROTEIN	838	t--taegtggaagatttgq--cagaagatcctccggccatgggttaaacattggttaaat 84 L T W Q D L A E D P P P W V X P W I N
Sbjct:	gag polyprotein	878	tggtat--tggcaggacctogccaga--atccccaccctgggt---gccagcct--- 83 V V W Q D L A Q N P P P W V P A
Query:	VIRAL CORE PROTEIN	894	aaccaaagaagacc-aggtccccgaatcctggtctcttgaga---gaaaaacaacact~ 103 K P R K P C P R I L A L G E K N K H
Sbjct:	gag polyprotein	926	---cc-----gccaaagtcgctg--t-tgtctct--gatacccg-aagac---cagtc 99 S A K V A V V S D T R K F V
Query:	VIRAL CORE PROTEIN	949	-cggeega-aaaagtcgagc-cct---cttctctgtatctaccaccgag-atic-gaggagc 121 S A E K V E P S S S Y L P R D R G A
Sbjct:	gag polyprotein	967	ggcg-ggaggcca-tc-Agtctctccccgacccccatcta-cccg-gcaacagacga-c 113 A G R P S A P P R P P I Y P A T D D
Query:	VIRAL CORE PROTEIN	1001	--cgccgact-tggccggaaaccccaacctgttccccaccccttatacagcacagggt 139 A D L A G T P T C S P T P L S S T G
Sbjct:	gag polyprotein	1021	ttactctctctct---ctgaaccc-ac-g-ccccgcgc---tatcggc---ggc 131 L L L L S E P T P P P Y P A A

Blast Result

Page 3 of 16

Query: 1057 gctgtgaggggaacctct-gcccctc--c--t-gga--gc--tccgg----tg--g-t  
VIRAL CORE PROTEIN 157 C G E G T S A P F G A P V V  
||| | |||| ||||| | | ||| | | ||| | |  
Sbjct: 1065 actgcca--ccccctctggccctcaggcgatcggaccgctcaggccagatgcccgat  
gag polyprotein 146 L P P P L A P Q A I G P P S G Q M P D

Query: 1097 -ggagggga-cct--gctgccgg--gactcggagccggagagggcgc--acccggag--cg  
VIRAL CORE PROTEIN 171 E G P A A G T R S R R G A T P E R  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1123 agtagcgatcctgaggggcccagccg-ctggga-ccag-gagtcgcccgtgcccgcagtcga  
gag polyprotein 166 S S D P E G P A A G T R S R R A R S P

Query: 1148 gacagacgaga-tcgcg--at-attaccgctgcgacactatggccctc--ccat-g--  
VIRAL CORE PROTEIN 187 T D E I A I L P L R T Y G P P M  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1180 g-cagao-a-actcgggtcctgaot-coactgtg-a-ttttgcctcgcgagccatagga  
gag polyprotein 184 A D N S G P D S T V I L P L R A I G

Query: 1197 --ccagg--g-ggccaattgc--agcccctc--cagattggcccttttcttctgcagat  
VIRAL CORE PROTEIN 204 P G G Q L Q P L Q Y W P F S S A D  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1234 ccccgccgagcccaatggcctggcctcctacatattggccttttcttctgcagat  
gag polyprotein 202 P P A E P N G L V P L Q Y W P P S S A D

Query: 1248 ctctataattggaaaactaaccatccccctttctcggaggatcccca--acgcctcagcg  
VIRAL CORE PROTEIN 221 L Y N W K T N H P P F S E D P Q R L T  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1294 ctttataattggaaatctaatccctcttttctga--aaacccagcaggtctcagcg  
gag polyprotein 222 L Y N W K S N H P S F S E N P A G L T

Query: 1306 ggttggtggagtccttatgttctctcaccagcctacttgggatgattgtcaacagctgc  
VIRAL CORE PROTEIN 240 G L V E S L M F S H Q P T W D D C Q Q L  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1352 ggctccttgagtccttatgttctcctccatcagccacttgggacgattggcaccagctcc  
gag polyprotein 241 G L L E S L M F S H Q P T W D D C Q Q L

Query: 1366 tgcagacactcttcacaaccgaggagcagagagagattctgttagaggctagaaaaaatg  
VIRAL CORE PROTEIN 260 L Q T L F T T E E R E R I L L E A R K N  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1412 tacagattcttttaccactgaggaacgggaaagaattctcctggaggcccgcaaaaatg  
gag polyprotein 261 L Q I L F T T E E R E R I L L E A R K N

Query: 1426 ttcc-tggggccgacggggcgaccc-acgcag-ttgcaaaa--tgagattgacatg-gg-a  
VIRAL CORE PROTEIN 280 V P G A D G R P T Q L Q N E I D M G  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1472 -tcttggggacaatggg-gccctacacagctcg-agaacct--catt-a-atgaggcc  
gag polyprotein 281 V L G D N G A P T Q L E N L I N E A

Query: 1479 ttcccttgactcg-cc-c-cggttgggactacaacacggcctgaaggtagggagag-ct-  
VIRAL CORE PROTEIN 298 P P L T R P G W D Y N T A E G R E S  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1525 ttccctcacaatcgacctcag---tgggattacaacacagcccgaggtagggagctctt  
gag polyprotein 299 F P L N R P H W D Y N T A A G R E R L

Query: 1534 -tgaaaatctatgccagg-ctctggtggcggtctc--cggggc-gcctcaagacggcc  
VIRAL CORE PROTEIN 316 L K I Y R Q A L V A G L R G A S R R P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1582 atg---gtctacggcc-ggaotctagtggtctctcaaaagggag-ctc--ggcgcc  
gag polyprotein 318 L V Y R R T L V A G L K G A A R R P

Query: 1589 cactaatttggctaaggtaagagaggtgatgcagggacc-gaacgaacctccc-tcggtg  
VIRAL CORE PROTEIN 335 T N L A K V R E V M Q G P N E P S V  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 1635 taccatttggctaaggtaagagaggtcttgcagggaccggca-gaac-cccttcggt-  
gag polyprotein 336 T N L A K V R E V L Q G P A E P P S V

Query: 1647 tttcttgaga--ggctcatggaagcct-tcaggcgggttcacccctcttgat-cctacctc  
<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?l> 21/02/01

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?>

21/02/01

blast Result

Page 5 of 16

Query: 2274 -ccattaggaacta---aaagnnnnnnntcctgggtgatgggtgccacagg--gca  
REVERSE TRANSCRIPTASE 38 P L G K L K E K K S W V M G A T G Q  
||| | ||||| | | | | | | | | | | | | | | | | |  
Sbjct: 2305 ccca-tgggaaaagtaggtccagacggacggtcggtg---ga-aggagcgacaggcagca  
pol polyprotein 38 P M G K V G S R R T V V E G A T G S

Query: 2327 acgg-cagtatccatggactaccgaagaaccgrr--gacttggga-gtgggacgggtaa  
REVERSE TRANSCRIPTASE 55 R Q Y P W T T R R T V D L G V G R V  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2360 a-ggtc--taccctggaccacaaaag-acttttaaaatt-ggacat-aaacaagtga  
pol polyprotein 56 K V Y P W T T K R L L K I G H K Q V

Query: 2383 cccactcgtttcttggtcat-ccctgagtgccag-taccct-tctaggtagagacttac  
REVERSE TRANSCRIPTASE 74 T H S F L V I P E C P V P L L G R D L  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2414 cccaotcctcctgtctacac-gagtgcctgtgt-oototgt-tgggaagggaotc  
pol polyprotein 74 T H S F L V I P E C P A P L L G R D L

Query: 2440 tgaccaaga-t-gggagctcaatttc--tttt--gaac-aaggaagaccagaagtgc-  
REVERSE TRANSCRIPTASE 93 L T K M G A Q I S F E Q G R P E V S  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2471 taaccaa-actaaag-gcccaga-tccagtttccg--ctgagg--gcccacaggttaaca  
pol polyprotein 93 L T K L K A Q I Q F S A E G P Q V T

Query: 2492 tgrgaataacaaccatcactgtgtgacc--ctcc-aa--ttagatgargaatatcga  
REVERSE TRANSCRIPTASE 111 V N N K P I T V L T L Q L D D E Y R  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2524 tggggagaac--gccc-t-actatg-tg-cctgggtcctaaacctggaagaagaatccga  
pol polyprotein 111 W G E R P T H C L V L N L E E E Y R

Query: 2547 ctatat-tctcccaagttaagcctgatcaagatat--a--cagtcctggttgagcagt  
REVERSE TRANSCRIPTASE 129 L Y S P Q V K P D Q D I Q S W L R Q  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2578 ctacatgaaaagc-agt--accctcctc---tatcgaccca-tcctgggtccagc-tt  
pol polyprotein 129 L H E K F V P S S I D F S W L Q L

Query: 2602 ttcccca-agcctgggcagaaaccgcagggtgggtttggcaagcaagttccccacag  
REVERSE TRANSCRIPTASE 147 F P Q A W A E T A G M G L A K Q V P P Q  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2629 ttcccaactgtatgggcagaaagagccggcatgggactagccaatcaag-tccac-cag  
pol polyprotein 146 F P T V W A E R A G M G L A N Q V P P

Query: 2661 -gttatthca-a-ctgaaggcca-gtg-ctacaccagtatcag-teagacagtacccttg  
REVERSE TRANSCRIPTASE 167 V I Q L K A S A T P V S V R Q Y P L  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2687 tggtagt-agagct-aagatcaggtgcct-caccagtggtgttc-gacaatatccaatg  
pol polyprotein 165 V V V E L R S G A S P V A V R Q Y P M

Query: 2715 agtagagaggctcagagaaggatttggccgcatgttca-aagattaatcc-a-ac--agg  
REVERSE TRANSCRIPTASE 185 S R E A R E G I W P H V Q R L I Q Q  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2743 agcaagaagctcgggaagggtatcagacccacatccagaag-r---tccagacctagg  
pol polyprotein 184 S K E A R E G I R P H I Q K F L D L G

Query: 2770 gcatacctagt-tcctgtccaatc-cccttgggaatactccc-ctgctaccggttaggaagc  
REVERSE TRANSCRIPTASE 203 G I L V P V Q S P W N T P L L P V R K  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2799 g-gtcttgggtgcctgt-cggtcgccc-tgggaatac-ccototgtaccctgtaaaaaagc  
pol polyprotein 202 V L V P C R S P W N T P L L P V K K

Query: 2827 ctgggaccaatgattatcgaccagttacaggacttgagagaggtcaataaaagggtgcagg  
REVERSE TRANSCRIPTASE 222 P G T N D Y R P V Q D L R E V N K R V Q  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 2855 cagggaccaatgactatcgccagttcaagacctgagagaaatataaaagggttacagg  
pol polyprotein 221 P G T N D Y R P V Q D L R E I N K R V Q

Query: 2887 acataca-cccaacggtcccgaaacattatac-ccctcttgagcgcctcctcgaagc  
http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi? 21/02/01

Page 6 of 16

21/02/01

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?>

21/02/01

### Blast Resist

21/02/01



pol polyprotein 843 S A V R E V T S Q C Q A C A M T N A V

Query: 4821 aatc-cttccaga-afacotccaggaagagactaa-gggga-aggcccccaggcgt-c  
REVERSE TRANSCRIPTASE 887 N P S R I P F G K R L A G S H F G A  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 4776 caccacctacagagagacc-----ggaaaaagggc-aacgaggagatcgacc-ggag-tgt  
pol polyprotein 862 T T Y R E T G K R Q R G D R P G V

Query: 4876 actgggaagtggactccactgaggttaagcc-ggctaatacgggaacaaatatctattg  
REVERSE TRANSCRIPTASE 905 H W E V D F T E V K P A K Y G N K Y L L  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 4829 actgggaggttagacttcacagaaataaagccctgg-tcggtatggaaacaaagtatctgtta  
pol polyprotein 879 Y W E V D F T E I K P G R Y G N K Y L L

Query: 4935 gttttctagacaccttt-tcagatagggtagaggcttatcctactaa-----gaaa-gag  
REVERSE TRANSCRIPTASE 925 V F V D T F S G W V E A Y P T K K E  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 4888 gtattcatagata-ctttctcgggatgggtagaagcttttccatccaaactgaaacg-g  
pol polyprotein 899 V F I D T F S G W V E A F P T K T E T

Query: 4939 acrtcaacogtgggtggctaagaaataactggaggaaattttccaagatttggaaaccc  
REVERSE TRANSCRIPTASE 943 T S T V V A K K I L E E I F P R F G I P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 4946 cccct-aatcgt-ctg--taaaaaaataattagaagaattctaccccgcttcgggatccct  
pol polyprotein 918 A L I V C K K Y L E E I L P R F G I P

Query: 5049 aaggtaataggggtcagacaatgggtccag-cttctggtgcccaggtaagtcagggtactggc  
REVERSE TRANSCRIPTASE 963 K V I G S D N G P A F V A Q V S Q G L A  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5002 aaggtactcgggtcagacaatgggtcccgcttt-gttgctcaggtaagtcagggtactggc  
pol polyprotein 937 K V L G S D N G P A F V A Q V S Q G L A

Query: 5108 ca-agatattggggattgattggaacatgcatgcatagacaccccaagctcaggac  
REVERSE TRANSCRIPTASE 983 K I L C I D W K L H C A Y R P Q S S G  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5061 cactca-actggggataaattggaagttacattgtgctatagacccagagctcaggtc  
pol polyprotein 957 T Q L G I N W K L H C A Y R P Q S S S

Query: 5167 aggtagagaggatgaatagaaccatttaagagaccctt-accnaatt-gaccacagagac  
REVERSE TRANSCRIPTASE 1002 Q V E R M N R T I K E T L T K L T T E T  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5120 aggtagaagaatgaacagaaataaagaga-ccttgacaaattag-ccttagagac  
pol polyprotein 976 Q V E R M N R T I K E T L T K L A L E T

Query: 5225 ---tgccattaatgattggatggctctcctgccccttttgcttttttaggggtgaggaacac  
REVERSE TRANSCRIPTASE 1022 G I N D W M A L L P F V L F R V R N T  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5178 cgggtg-a--aaagactgggtgaccctccttcccttagcgtgcttagggccagggaatac  
pol polyprotein 996 G G K D W V T L L P L A L L R A R N T

Query: 5282 ccctggacagtttggtgacccctatg-aattgctctacggggga-----nnnnnnhng  
REVERSE TRANSCRIPTASE 1041 P G Q F G L T P Y E L L Y G G P P P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6235 ccctggcgggtttggttaactccttatgaaatt-ctctatggaggaccacccccatac  
pol polyprotein 1015 P G R F G L T P Y E I L Y G G P P P I

Query: 5337 ttg-g-c---agaaattgccccttgacatagtgctgatg-t-gctgctt-tcccagcctt  
REVERSE TRANSCRIPTASE 1059 L A E I A F A H S A D V L L S Q P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6294 ttgagtgtggagaaa-----ctttg-----ggtooggtgatagat--ttctcctgtctt  
pol polyprotein 1034 L E S G E T L G P D D R F L P V L

Query: 5389 -tgctt-ctctaggc-tcaagggcgtcg-agtgggtgaggcagc--gagcgtggaagcagc  
REVERSE TRANSCRIPTASE 1076 L F S R L K A L E W V R Q R A W K Q  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5343 at-ttactc-a--cttaaggccttagaaat-tgtaagg-accnaaact-tgggaccaga  
pol polyprotein 1050 F T H L K A L E I V R T Q I W D Q

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>

21/02/01

### Blast Result

```
Query:          5443 tccgggaggcctactacg-gaggagac--ttgcaggttccaca-tcgcttccaagtggga
REVERSE TRANSCRIPTASE 1094 L R E A Y S G G D L Q V P H R F Q V G
                  ||| | | | | | | | | | | | | | | | | |
Sbjct:          5396 tcaaaagggtgtg-taacgctgg-taccgttaacaa-tccctcacceg-ttcagggtcggg
pol polyprotein 1058 I K E V Y K D C T V T I P H P F Q V G
```

```

Query:          5499 gattc-agt-ctatgttagacgccaccg-tgcaggaacctcgagactcgggtggaaggga
REVERSE TRANSCRIPTASE 1113 D S V Y V R R H R A G N L E T R W K G
                  || || | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:          5452 ga-tcaagtgtct-tgtcagacgccatcgaccca-gcaqcccttgagcctcgggtggaaggc
pol polyprotein 1067 D Q V L V R R H R P S S L E P R W K G

```

```
Query:          5556 ccttatctcgtgaactt-ttgaccacaccaacggcgctgtaagaagtgcgaaggaaatcccacctg
REVERSE TRANSCRIPTASE 1132 P Y L V L L T T F T A V K V B G I P T W
               ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:          5509 ccataccctggg-gttgctgactaccccgaccgcggtaaaagtcgatgtattgtgttcgctg
pol polyprotein   1106 P Y L V L L T T F T A V K V D G I A A W
env protein       1                               M V L L P
```

```

Query:                5615 gatccatgcatccccagcttaagccggcgccacctcccg-actc-g---ggg---tggagag
VIRAL ENVELOPE PROTEIN 1      M H P T L S R A R P H L P T S R G C R
REVERSE TRANSCRIPTASE 1152   I H A S H V L K P A P P D S G W R
                             | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:                5568 ggtccatgctctctcacctcaaacctgcaccaccttcggcac-cagatgagtcctgg-gag
env protein           6      G S M L L T S N L H H L R R H Q M S P G S
pol polyprotein       1126   V H A S H L K P A P P S A P D E S W E

```

```

Query:          6668 c-cgaaaagaactgagaaatccccctaagcttcgcctcca-cc-gc--ctggttccttac--
VIRAL ENVELOPE PROTEIN 17  P K R L R I P L S F A S I A W F L T
REVERSE TRANSCRIPTASE 1169 A E K T E N P L K L R L H R L V P Y
                        | | | | | | | | | | | | | | | | | | | | | |
Sbjct:          5626 ctggaaaagaactgatacatccrcttaagct--gcgt--attcgccggc-gg--cgggacga
env protein     25  W K R L I I L L S C V F G G G G T
pol polypeptin 1145 L E K T D H P L L K L R I R R R R P E

```

```

Query:          5721 -tct--aacaat---aactcc--ccaggccagtagtaaacgcctta-tagacagct-c-g
VIRAL ENVELOPE PROTEIN 35      L   T   I       T   P   Q   A   S   S   K   R   L   I   D   S   S
REVERSE TRANSCRIPTASE 1187    s   N   N       N   S   P   G   Q   ^^^
                        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Sbjct:          5679 gtcctgcaa-aataaggaacccccaccagccca-t-g---ac-cctcacttggcagggtactg
env protein     43      S   L   Q   N   K   N   P   H   Q   P   M       T   L   T   W   Q   V   L
pol polypeptide 1162    S   A   K   ^^^

```

```

Query:          5770 aacccccca-t--agacott-tatccotttac-ctggctgat-tattgaccc--tgatac-g
VIRAL ENVELOPE PROTEIN 51  N P H      R P L S L T W L I I D P D T
                  ||| | | | | | | | | | | | | | |
Sbjct:          5732 -tcccaaatctggagagctgtgtctgggatacaaaagcgag-tcca--gcccccttg-gactt
env protein     61  S Q T E D V V W D T K A V Q P P W T

```

```

Query:          5821 ggtg----tcactgtaaa--t-a-gca---c--tcgagg-t-g--ttg-cttct-aga-
VIRAL ENVELOPE PROTEIN 68  G      V  T  V  N      S      T      R  G      V  A  P  R
                    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Sbjct:          5787 ggtggcccaact--taaaccgtatgtatgtgccttgccggctagtctctgagtcctcgagat
env protein      79  W  W  P  T  L  K  P  D  V  C  A  L  A  A  S  L  E  S  W  D

```

```

Query:          5860 -----ggcacc--tg-gt-ggc--ctgaactgcatttctg-cctcc-gatt--gatta-
VIRAL ENVELOPE PROTEIN 81      S T W W P E L H F C L R L I
               || ||| || | | | || ||| || | | || ||
Sbjct:          5846 atcccgggaaccgatgtctcgtcctctaaac-g-a-gtcagacctccggactcagactat
env protein     99      I F G T D V S S S K R V R F F D S D Y

```

```

Query:          5903 ac-cccgtgtgtaaaagca--cacct-----ccca-a-----c--ctagtc-cg--ta
VIRAL ENVELOPE PROTEIN 95  N  P  A  V  K  S      T  P      P  N      L  V  R
                  || |||| ||| ||||| ||||| ||||| ||| ||| ||| |||
Sbjct:          5903 actgcgcg--ctataagcaaaacacctggggagcctatagggtgcagctaccctcggggcta
env protein     118  T  A  A      Y  K  Q  I  T  W  G  A  I  G  C  S  Y  P  R  A

```

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?>

21/02/01

BLAST RESULT

Page 11 of 16

Query: 5942 g---t--tatgg---gttcta--tt---g-ctg-ccc---a-ggc---ac---agag  
VIRAL ENVELOPE PROTEIN 108 S Y G F Y C C P G T E  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 5961 ggactagaatggcaagctctacettctacgtatgtcccggtatggccggaccctttcag  
env protein 137 R T R M A S S T F Y V C P R D G R T L S

Query: 5974 aaag--agaa-atactgtgggggttctggggaat-ect-t-ctgt-aggagat-gga--g  
VIRAL ENVELOPE PROTEIN 119 K E K Y C G G S G E S F C R R W S  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6021 -aagctagaaggtgc-g-ggggg--ct--agaatccctatactgtaaaga-atgggattg  
env protein 157 E A R R C G G L S S L Y C K E W D C

Query: 6024 ctgcgtcacctccaacgatggagactgg-aaatggccgactc--ct-ctc-caggaccgg  
VIRAL ENVELOPE PROTEIN 135 C V T S N D G D W K W P I S L Q D R  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6073 -rdag--acc---acg--gg-gacgggttattgg-ctatcctaactcctcaaaagacc--  
env protein 175 E T T G T G Y W L S K S S K D

Query: 6079 gt-aaaattctcctttgtcaattccggccggggaagta-aaaatgagaaaot-atat  
VIRAL ENVELOPE PROTEIN 184 V K F S P V N S G P G K Y K M M K L Y  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6120 -tcataa--c-----tgtaaaat-gggacc--aaaatagc-gaatg--g--actca-a-  
env protein 190 L I T V K W D Q N S E W T Q

Query: 6136 aaagat-aagagctgctcccca-tcagacttagat-tatc-----taaagata-agtttc  
VIRAL ENVELOPE PROTEIN 173 K D K S C S P S D L D Y L K I S F  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6161 aaatttcaacag-tg-tcaccagaccggct--ggtgtaaacccctttaa-aataga-tttc  
env protein 204 K F Q Q C H Q T G W C N P L K I D F

Query: 6187 actgaaaggaaaacaggaaaa-tattcaa-aagtggataaatgggtatgagctgggggaat-  
VIRAL ENVELOPE PROTEIN 190 T E R K T G K Y S K V D K W Y E L G N  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6215 ac--a---gacaa-aggaaaattatccaaggactggataacgggaaaaaacctgggggarta  
env protein 222 T D K G K L S K D W I T G K T W G L

Query: 6244 agttttttattatattg-g-cggggga-gcagg-gtccactttaaccattcgcttaggat  
VIRAL ENVELOPE PROTEIN 209 S F L L Y G G G A G S T L T I R L R I  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6269 ag-----attctatgtgtcttgacatccaggcgtaca-gttcaccattcg-ctta--a-  
env protein 240 R F Y V S G H P G V Q F T I R L

Query: 6300 aga-gacgggggaca-gaacccctgtggcaatgggaccgga--taaagtactggct-gaa  
VIRAL ENVELOPE PROTEIN 228 E T G T E P P V A M G P D K V L A E  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6318 aaatcac--caacatg--ccagctgtggcagtaggtcctgacct--cgctcttg-tggaa  
env protein 256 K I T N M P A V A V G P D L V L V E

Query: 6355 caggggcc-cc-----cggccctgg---agccaccgcataacttggc---gg--tgccc  
VIRAL ENVELOPE PROTEIN 246 Q G P P A L E P P H N L P V P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6371 caaggacctcctagaacgtccctcgtctccacactcct--cttcccccaagggaagcgc  
env protein 274 Q G P P R T S L A L P P P L P P R E A

Query: 6400 caattaacc-tcgtcgggcctgac-ataa---cacagcc---gc--cta--gca-acag  
VIRAL ENVELOPE PROTEIN 261 Q L T S L R P D I T Q P F S N S  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6429 ca--ccgcatctct-c---cccgactctaactccacagcctggcgactagtcacaaac  
env protein 293 P F P S L P D S N S T A L A T S A Q T

Query: 6447 taccactggattgattcctaccaa---caagcctagaa-actcc--ccaggtgttccctgt  
VIRAL ENVELOPE PROTEIN 277 T T G L I P T N T P R N S P G V P V  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6484 tccac-gg--tga-gaaaaacaattgttac-ccta-aacactccgccc-----tcc--  
env protein 312 P T V R K T I V T L N T P P P

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?1>

21/02/01

## Blast Result

Page 12 of 16

Query: 6501 taagacaggac-agagac-tcttcagttctcatccaggagctttcc-aagccatcaact-  
VIRAL ENVELOPE PROTEIN 295 K T G Q R L F S L I Q G A F Q A I N  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6529 caccacagg-cgacagactttttga-tcttctgagggggccttctaa-cc-ttaaatg  
env protein 327 T T G D R L F D L V Q G A F L T L N

Query: 6557 ccaccgaccctgatgccacttcttctgttggtttgtctat-cctca-ggg-cctcctt  
VIRAL ENVELOPE PROTEIN 313 S T D P D A T S S C W L C L S S G P P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6585 ctaccaacccaggggccactgagttctgttggtttgt-t-tggc-catggggccctt  
env protein 345 A T N P G A T E S C W L C L A M G P P

Query: 6614 attatgaggggat-ggctaa--agaaagaaaattcaatg--tga--ccaaagagcataga  
VIRAL ENVELOPE PROTEIN 332 Y Y E G M A K E R K F N V T K E H R  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6642 attatgaaggaatagcttcacag--gagaggto--gect-actccacaga-ccttg-  
env protein 364 Y Y E A I A S S G E V A Y S T D L

Query: 6667 aatcaatgtac-atgggggtcccgaa--ataagcttacccctcactgaagttccggga-a  
VIRAL ENVELOPE PROTEIN 350 N Q C T W G S R N K L T L T E V S G K  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6693 -accggtg-cgctgtgggggaccc-aagga-aagctcaccctcactgaggtct-caggaca  
env protein 381 D R C R W G T Q G K L T L T E V S G H

Query: 6723 ggggacatgcctaggganaagctcccccac-ccccccccaccccttgctat-agtaactgtg  
VIRAL ENVELOPE PROTEIN 369 G T C I G K A P P S H Q H L C Y S T V  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6748 cggtgtgtgcatagg-aaggtgcccctttaccatcagcatctctgaatcag-accct-  
env protein 400 G L C I G K V P F T H Q H L C N Q T L

Query: 6781 gtttatgagcaggcctc--agaaaatcagtatttag-tacctggttataac--aggtg  
VIRAL ENVELOPE PROTEIN 388 V Y E Q A S E N Q Y L V P G Y N R W  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6805 atccatca--attcctccggagaccatcagtatct-gctccc--ctccaaccatagctg  
env protein 419 S I N S S G D H Q Y L L P S N H S W

Query: 6834 gtggcatgcaataactgggttaacccctgtgtttccacctcagttctcaacca-atcca  
VIRAL ENVELOPE PROTEIN 406 W A C N T G L T P C V S T S V F N Q S  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6869 gtgggttgagcactggcctcacccttgccctccacctcagtttttaatacaga-cta  
env protein 437 W A C S T G L T P C L S T S V F N Q T

Query: 6893 aagattttctgtgt-catgggtccaaatcg-tcccccg-agtgactaccatcctgaggaag  
VIRAL ENVELOPE PROTEIN 425 K D F C V M V Q I V P R V Y H P E E  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6918 gagatttctgtatcca-ggtccagct-gattcctcgca-tctattactcctgaagaag  
env protein 456 R D F C I Q V Q L I P R I Y Y F E E  
CLEARANCE

Query: 6950 tggctccttgatga-a--tatgactatcggtataacc--gaccaaagagagacccg  
VIRAL ENVELOPE PROTEIN 444 V V L D E Y D Y R Y N R P K R E P  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 6975 t--t--ttg-ctacaggcctatgacaat--tctcaccacagga-ctaaaagagaggctg  
env protein 475 V L L Q A Y D N S H P R T K R E A

Query: 7001 tacccttaccctagctgtatgtctcggattaggg----acggccgttgccgttaggaaca  
VIRAL ENVELOPE PROTEIN 461 V S L T L A V H L G L G T A V G V G T  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 7026 tctcacttaccctagctgttttactgggtt-gggaaacacgg-cg--gggaataggtact  
env protein 492 V S L T L A V L L G L G I T A G I G T

Query: 7057 gggacagctgccctgatcagagacc-acagcagctagagaagggacttggtgag-ct--  
VIRAL ENVELOPE PROTEIN 480 G T A A L I T G P Q Q L E K G L G E L  
| | | | | | | | | | | | | | | | | | | | | |  
Sbjct: 7082 ggttcaactgccttaattaaaggacatag-accctcagcaagg-cctgacaagcctcc  
env protein 511 G S T A L I K G P I D L Q Q G L T S L

Query: 7113 acatgoggocat-gaaagaagatctccgag-cottaaaggagctgtttagcaacctagaa  
http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?l 21/02/01

Query: 7791 gta-aga---taacaggaaatgagttgact-aatc-gcttatctggatt-ctgaa-aa

21/02/01

## Blast Result

Sbjct: 7790 ggacagaggggtccca-gaaat-agatgagtcac-cagc--a----gtttccagcaagga  
Query: 7842 ctgactgg--caccataga-agaattga-ttacacattgacag--cccta--gtgaccta  
Sbjct: 7841 caga-ggggttc-ccagaaataga--tgagtca-aca--g-cagtttcc-aggggtgccc--  
Query: 7894 tctcaactgcaatctgtcactctgccagggagccacgcagatgctggacacctccggagcta  
Sbjct: 7890 -ctcaaccg---t-t-tca-----agga-ctc-c-c-at---ga---ccgg-g---  
Query: 7954 ttttaaaatgattggtccacggagcgggctctcgatattttaaaatgattggtcc-a-t  
Sbjct: 7921 -----aat-t-----cac-----ccctggc-ct---ta-tttgaactaa-----ccaat  
Query: 8013 ggagcgc-gggctctcgatattttaaaatgattggttctgacgcacag-gctttgttg-  
Sbjct: 7954 -ta-c-cttgccctctcgct-----tctgt-ac-c-c-gcgttt-ttg  
Query: 8070 tga-accocataa-aagctgtcccgattc-cgcaactcggggc-cgcagtcctctacccct  
Sbjct: 7990 t-ata-----a-ataagct---cagaaactc-caccggagcgc-cagtcct-ta-----  
Query: 8126 gcgtg-gtcta--cgactgtgggcccagggcgttggaaataaaatcctctgtgttt  
Sbjct: 8033 gagagactg-agccgccc-ggtaccc-gtgtg--tccaataaaa--cctctgtgtgatt  
Query: 8183 gca 8185  
Sbjct: 8086 gca 8088

Score = 113 bits (1449), Expect = 9e-22  
Identities = 577/1182 (48%), Gaps = 460/1182 (38%)  
Strand = Plus / Plus

Query: 7415 ccc---tgetttctg-ctctga--cg---ggg---ccoctagtagtcc--t---gtc--  
VIRAL ENVELOPE PROTEIN 599 T L L S A L T G P L V V L L  
Sbjct: 2 cgcagtccttagagagactgagccgcccgggtaccggt-gt-gtccaataaaacctctt  
Query: 7456 -ctg-ttacttac--ag--ttgggc-c-ttg---ctt---a--at-----taa-ta--g  
VIRAL ENVELOPE PROTEIN 613 L L L T V G P C L I N R  
Sbjct: 60 gctgattgcacccggaqccgtggtctcgttgccttgggaggggtttctcctaactattg  
Query: 7491 ---g---tt---tgt-tgc-c-tttg---t--t---aga---gaa---cgagtga  
VIRAL ENVELOPE PROTEIN 624 F V A P V R E R V  
Sbjct: 120 accgcccacttcgggggtct-cacatttgggggtcgtccgggatcggaaacccaccca  
Query: 7520 gtg--ca--g-tcca-gatc---atggta--cttaggcaacag-tacc-aag-g---c--  
VIRAL ENVELOPE PROTEIN 634 S A V Q I M V L R Q Q Y Q G  
Sbjct: 178 gggacccagagcccaacggga-dgtaagct--dgc--cagcagaccgtgtgtgtctc  
Query: 7561 -cctctgag-cgaag-gagaaa--atgac-ct-atagcct-tc-c---cagt-tct-a-  
VIRAL ENVELOPE PROTEIN 648 L L S Q G E T D L ^^^  
Sbjct: 234 gcttctgtgtcctaagtcgtaattctgactgtccttgtgtgtctcgtctctgtgtgag  
Query: 7605 a--g-a--t-t-a--gaactat-taa---caagaca-----ag---aagt-----g  
Sbjct: 294 aocgtaactctgactgcoot-tgtaagtgcgcg-catttttttgggttcaactctattccg

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?1>

21/02/01

### Blast Result

Query: 7635 gg-gaa-----tgaaag-ga--tgaaat-gc-aacct-a-c-----ctc-  
 Sbjct: 352 ggtgaatcactctgcgagtgacgtgtgagtagcgaaac--agacgtgttcggggctcaccg

Query: 7669 ccag--aaccca-gga-a-gt--ta--at-a---a-a-a---a-gctctaaatg--c  
 Sbjct: 410 cctggtaatccaggagacgtcccaggatcaggggagggaccaggagacgc-ct-ggtggac

Query: 7705 ccc-cg--aa----t--tacagaccctgctggc-t-gccagtaaataggt-agaag-g-t  
 Sbjct: 468 ccctcggtaacgggtcgttctgaccc-qatttcatcgccgt--ct-ggtaagacgcgt

Query: 7751 cacacttctt-attgt-te-c-agg--gact---gacac-ctgg---c-ctaag-taa  
 Sbjct: 524 ctgaat--ctgattctctctctcgtcgcctcgccgctctctggttctctctctctc

Query: 7795 gataacaggaaa---tgagt-tgactaatc--gct-t-----atc-tggat-t----c-t  
 Sbjct: 582 g-tttctggaaagcctctgtgtcac-agtcttctctcccaaatcatcaatatgggacaa  
 gag polyprotein 1 M G Q

Query: 7836 g-taa----a--actga-ct--ggcac-c---at-a--gaaga-at-tgattacaca---  
 Sbjct: 640 gataattctaccct-atctccctcactctaaatcactggagagatgtgagaaca-aggg  
 gag polyprotein 4 D N S T F I S L T L N H W R D V R T R

Query: 7874 ttgacagccota---gt-gacctatctcaactgcaatctgtc--ac--tctg--cc-c--  
 Sbjct: 698 ctccca-atctatccgtgga--aaccaaaaagggaataatggcagactttctgttccctccg  
 gag polyprotein 23 A H N L S V E I K K G K W Q T F C S S

Query: 7921 ag-gagccac-----gcagatgcg--gacctccggag---ct-----a  
 Sbjct: 755 agtg-gcccacattcggc-g-tgggggtggccacgggaggaacttttaattctctgtca  
 gag polyprotein 42 E W P T F G V G W P P E G T F N L S V

Query: 7954 ct-----ttaaatgattggtc---cacggag--cgcggg-c-tctc-ga--tatttt  
 Sbjct: 812 tttttgcagttaaaagatt-gtctttca-ggagaacgggggacatc-cggaccaagttc  
 gag polyprotein 61 I F A V K K I V F Q E N G G H P D Q V

Query: 7997 aaaatgat--tggtccatgg-ag--cgcggtctctc-gatat--tttaaatgattgggt-  
 Sbjct: 869 catat-atcgtggt--atggcaggac-ctcgc-c-caga-atccccacacatg---ggtg  
 gag polyprotein 80 P Y I V V W Q D L A Q N P P P W V

Query: 8048 -ttg--tgacg-cacag---gct-ttgt-tgaga-aac-----aca-t-----a---  
 Sbjct: 919 ccagcct-cggcca-aggctcgtgtgtctctgataccggaagaccagttgcyggggaggc  
 gag polyprotein 97 P A S A K V A V V S D T R R P V A G R

Query: 8081 -a-aagct-gt-cccgattccgca-ct---cgg--v-ggccc-g--agtct-ct-----  
 Sbjct: 977 catcagcttctcccggaaccccaatctacccgggaacagacgacttactctctctctctg  
 gag polyprotein 116 P S A P P R P P I Y P A T D D L L L L S

Query: 8120 -accc-ctg---cg---t---gg-tgtac-g--a---ctgtggggccc-cagcgcgct  
 Sbjct: 1037 aaccac-gccccgcctctatcggcgccactggcaccctctgtggccctccag-gcgat  
 gag polyprotein 135 E P T P P P Y P A A L P P P L A P Q A I

Query: 8158 tggg-----ataa-----aat-cctcttgct-gtttgc-ata 8187

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?1>

21/02/01

## Blast Result

Subject: 1095 cggaccgcccgtcaggccagatgcc-c--gatag-taggatc 1132  
gag polyprotein 156 G P P S G Q M P D S S D

CPU time: 6.79 user secs.

0.19 sys. secs

6.98 total secs.

Gapped  
Lambda K H  
0.0514 0.0120 0.0135

Gapped  
Lambda K H  
0.0514 0.0120 0.0135

Matrix: blastn matrix:5 -2  
Gap Penalties: Existence: 1, Extension: 2  
Number of Hits to DB: 7964  
Number of Sequences: 0  
Number of extensions: 7964  
Number of successful extensions: 13  
Number of sequences better than 10.0: 1  
length of query: 8209  
length of database: 2,812,937,656  
effective HSP length: 1920  
effective length of query: 6289  
effective length of database: 2,812,935,736  
effective search space: 17690552843704  
effective search space used: 17690552843704  
T: 0  
A: 0  
X1: 135 (10.0 bits)  
X2: 1347 (99.9 bits)  
S1: 250 (24.9 bits)  
S2: 463 (40.7 bits)

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?l>

21/02/01



Blast Result

Attachment B



## Blast 2 Sequences results

PubMed

Entrez

BLAST

OMIM

Taxonomy

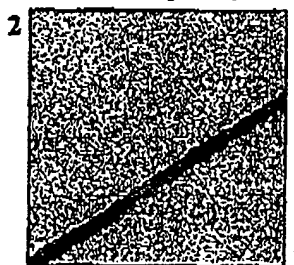
Structure

## BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.1.2 [Nov-13-2000]

Match:  Mismatch:  gap open:  gap extension: x\_dropoff:  expect:  wordsize:  Filter ☐

Sequence 1 gi 4538106 Sequence 3 from Patent WO9740167. Length 1575 (588 .. 2162)

Sequence 2 gi 9630311 Gibbon leukemia virus, complete genome Length 2370 (631 .. 3000)



NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

NOTE: If protein translation is reversed, please repeat the search with reverse strand of the query sequence

Score = 362 bits (4802), Expect = 2e-98  
 Identities = 1164/1749 (66%), Gaps = 345/1749 (19%)  
 Strand = Plus / Plus

Query: 588 atgggac-agacagtgaactaccccccttagt-t---tgactctcgaccattggact  
 VIRAL CORE PROTEIN 1 M G Q T V T T P L S L T L D H W T  
 ||||| ||| | | ||||| | | | | ||||| | | |||||  
 Sbjct: 631 atgggacaagataat-tctacccc---ta-tctccctcactctaaatcactgga--  
 gag polyprotein 1 M G Q D N S T P I S L T L N H W

Query: 642 -gttagatccaggggtcataattgtcagttcagg---ttaagaagggaccttggc  
 VIRAL CORE PROTEIN 19 V R S R A H N L S V Q V K K G P W  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct: 684 tgtgagaacaaggggtcacaatctatccgt---ggaaatcaaaaagggaaaatggc  
 gag polyprotein 18 V R T R A H N L S V E I K K G K W

Query: 698 tttctgtgocctctgaatggccaacattcgatgttgatggccatcagaggggaccc  
 VIRAL CORE PROTEIN 38 F C A S E W P T F D V G W P S E G T  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct: 741 tttctgttccctccagtggtccacattcggcgtgggtggccaccggagggaactt  
 gag polyprotein 38 F C S S E W P T F G V G W P P E G T

Query: 758 t---tctgaaattatccgtggtgtaaggcaatcatt---tttca-gactggacc  
 VIRAL CORE PROTEIN 57 S E I I L A V K A I I F Q T G P  
 | |||| ||| | | |||| | | ||| ||||| || |||  
 Sbjct: 801 tctctctgtcatt-t-ttgcagttaa---aaagattgtcttcagga---ga-ac  
<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0> 22/02/01

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0>

22/02/01

Attachment C



## Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.1.2 [Nov-13-2000]

x dropoff: 100 expect: 10.00 wordsize: 7 Filter ☐

Sequence 1 gi 4538106 Sequence 3 from Patent WO9740167. Length 3585 (2163 .. 5747)  
Sequence 2 gi 9630311 Gibbon leukemia virus, complete genome Length 8088 (1 .. 8088)

NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

NOTE: If protein translation is reversed, please repeat the search with reverse strand of the query sequence

Score = 798 bits (10671), Expect = 0.0  
Identities = 2598/3890 (66%), Gaps = 567/3890 (17%)  
Strand = Plus / Plus

```

Query:          2163 gggagacgggggttcggacccccctccccgagccagggtaactttgaaggtgg
REVERSE TRANSCRIPTASE
1             G R R G S D P L P E P R V T L K V
|||||
2194 gggagtccaggggttcggacccccctccccgaacctagggttaacactgactgtgg
sbjct:          1      G S Q G S D P L P E P R V T L T V
pol polyprotein

```

```

Query:          2222 accagttgagttcctggttgataccggagcggagcattcagtgctg-ctaca
REVERSE TRANSCRIPTASE 21      P V E F L V D T G A E H S V L L Q
                        ||| ||||| ||||| || ||||| || ||||| || |||||
Sbjct:          2254 ccca-ttgagttcctggtcgacacggagctgaacattcagttattgac-cca
pol polyprotein  21      P I E F L V D T G A E H S V L T Q

```

```

Query:          2280 ggaaaacta----aaagaaaaaaaatcctgg--gt-gatgggtgccacagg-
REVERSE TRANSCRIPTASE 40  G K L      K E K K S W      V M G A T G
                   ||||| ||      ||| |      || |||| || || |||||
Sbjct:          2311 ggaaaagttagggtccagacgga----c--ggtcgtgga--aggagcgacaggc
pol polyprotein  40  G K V G S R R      T      V V E G A T G

```

```

Query:          2331 -cagtatccatggactaccggaagaaccgtt---gacttggga-gtgggacg
REVERSE TRANSCRIPTASE 57  Q Y P W T T R R T V D L G V G R
                  | | | | | | | | | | | | | | | | | |
Sbjct:          2363 tc--taccctggaccacaaaaaga--cttttaaaaatt-ggacat-aaaca
pol polyprotein 57  V Y P W T T K R L L K I G H K Q

```

```
Query:          2386 actcgtttctgggcat-cctcgagtgcccag-taccctt-tctaggtagaga
REVERSE TRANSCRIPTASE 75 H S F L V I P E C P V P L L G R D
                  ||| || | | | | | | | | | | | | | | | | | | | |
                2417 actccttcctgggcatacc-gagtgcacctgcr-ctctgt-tgggcaggga
Sbjct:         75 H S F L V I P E C P A P L L G R D
pol polyprotein                               22/02/01
```

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0>

## Blast Result

Query: 2443 ccaaga-t---gggagctcaaatttc---tttt--gaac-aaggaagaccag  
REVERSE TRANSCRIPTASE 94 T K M G A Q I S F E Q G R P  
||||| | | | | | | | | | | | | | | | | |  
Sbjct: 2474 ccaa-actaaag---gcccaga--tccagttttccg--ctgagg--gccca-  
pol polyprotein 94 T K L K A Q I Q F S A E G P

Query: 2491 c-tgtgaataacaaacccatcactgtgtgaacc--ctcc-aa--ttagatga  
REVERSE TRANSCRIPTASE 110 S V N N K P I T V L T L Q L D D  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2522 catggggagaaac--gccc-t-actatg-tg-cctgggtcctaaacctggaaga  
pol polyprotein 110 T W G E R P T M C L V L N L E E

Query: 2545 gactatattctcccaagtaaagcctg-a---tcaagatat--a--cagtc  
REVERSE TRANSCRIPTASE 128 R L Y S P Q V K P D Q D I Q S  
||||| | | | | | | | | | | | | | | | | |  
Sbjct: 2576 gacta-----catgaaaagccagracccctc-ctctatcgaccca-tcc  
pol polyprotein 128 R L H E K P V P S S I D P S

Query: 2597 gcagtttcccca-agcctgggagaaaccgcag--gg-atgggtttggcaaa  
REVERSE TRANSCRIPTASE 146 Q F P Q A W A E T A G M G L A K  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2625 gc-ttttccccactgtatgggcagaaa--g-agccggcatgggactagccaa  
pol polyprotein 145 L F P T V W A E R A G M G L A N

Query: 2653 cccacag-gttattca-a-ctgaaggcca-gtg-ctacaccagtatcag-t  
REVERSE TRANSCRIPTASE 164 P P Q V I Q L K A S A T P V S V  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2680 ccac-cagtggtagt-agagct-aagatcaggtgcct-caccagtggctgtt  
pol polyprotein 163 P P V V V E L R S G A S P V A V

Query: 2707 accccttgagtagagaggtctcgagaaggaatttggccgcatgttca-aagat  
REVERSE TRANSCRIPTASE 182 Y P L S R E A R E G I W P H V Q R  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2735 atccaatgagcaaaagagctcggaaggtatcagacccacatccagaag-t  
pol polyprotein 181 Y P M S K E A R E G I R P H I Q K

Query: 2765 -ac--agggcatcctagt-tcctctccaatc-cccttgggaatactccc-ctg  
REVERSE TRANSCRIPTASE 202 Q G I L V P V Q S P W N T P L  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2791 gacctaggg-gtcttggtgcccctgt-cggtcgccc-tggaatac-ccctctg  
pol polyprotein 200 D L G V L V P C R S P W N T P L

Query: 2819 caggaagcctgggaccaatgattatcgaccagtacaggacttgagagaggtc  
REVERSE TRANSCRIPTASE 220 R K P G T N D Y R P V Q D L R E V  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2847 aaaaaagccagggaccaatgactatcgccaggttcaagacctgagagaatt  
pol polyprotein 219 K K P G T N D Y R P V Q D L R E I

Query: 2879 ggtgcaggacataca-cccaacgggtcccgaacccttataa-cctcttgagcg  
REVERSE TRANSCRIPTASE 240 V Q D I H P T V P N P Y N L L S  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2907 ggtacaggatatttcacccc-acagtcaccaaaccccttacaattctc-tgagtt  
pol polyprotein 239 V Q D I H P T V P N P Y N L L S

Query: 2937 cctgaacgga-ac-tggtacacagtattgga-cttaaaagatgccttcttct  
REVERSE TRANSCRIPTASE 259 P E R N W Y T V L D L K D A F F  
| | | | | | | | | | | | | | | | | |  
Sbjct: 2965 cct-agc-tatacttggtactcagtccttagatc-tcaaggatgccttttct  
22/02/01  
http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0

### Blast Result

Page 1 of 5

## Attachment



## Blast 2 Sequences results

PubMed

**Entrez**

**BLAST**

**OMIM**

## Taxonomy

## Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.1.2 [Nov-13-2000]**

Match:  Mismatch:  gap open:  gap extension:

x dropoff: 100 expect: 10.00 wordsize: 7 Filter ☐

Sequence 1 gi\_4538106 Sequence 3 from Patent WO9740167. Length 1971 (5620 .. 7590)

Sequence 2 gi\_963031.1 Gibbon leukemia virus, complete genome Length 2449 (5552.. 8000)

**NOTE:**The statistics (bitscore and expect value) is calculated based on the size of nr database

**NOTE:** If protein translation is reversed, please repeat the search with reverse strand of the query sequence

Score = 410 bits (5448), Expect = e-112  
Identities = 1396/2290 (60%), Gaps = 560/2290 (24%)  
Strand = Plus / Plus

```

Query:                5620 atgcatccacggttaagccggcgccacctcccg-actc-g--ggg---tgga
VIRAL ENVELOPE PROTEIN 1   M H P T L S R R H L P T R G G
REVERSE TRANSCRIPTASE 1153 H A S H V K P A P P P D S G W
                        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct:                5573 atgctttctacctcaaacctgcaccaccttcggcac-cagatgagtcctctgg-
env protein           8   M L L T S N L H H L F H Q M S P G
pol polyprotein      1127 H A S H L K P A P P S A P D E S W

```

```

Query:                5672 aaagactgagaatccccttaagcttgcctcca-tc-gc--ctggttcotta
VIRAL ENVELOPE PROTEIN 18  K R L R I P L S F A S I A W F L
REVERSE TRANSCRIPTASE 1171  K T E N P L K L R L H R L V P Y
                        ||||| ||| ||||| || | | | | | | | | |
Sbjct:                5631 aaagactgacatcctctttaagct--gcgt--attcgcgcg-gg--cgga
env protein           27  K R L I I L L S C V F G G G G
pol polyprotein       1147 K T D H P L K L R I R R R R D

```

```

Query:          5724 -aacaat---aactcc--ccag-----g-c--ca-----gta--gt--
VIRAL ENVELOPE PROTEIN 36      T I      N S P      Q      A      S      S
REVERSE TRANSCRIPTASE 1188      N N      N S      P      G      Q      ^  ^
                        || |||  ||| ||  ||||  ||  ||  |||  ||
Sbjct:          5684 caa-aataagaaacccccaccagcccatgaccctcacttgcaggtactotcc
env protein     45      Q  N K N P H Q P M T L T W Q V L S
pol polyprotein 1164 A      K ^^^

```

```

Query:          5754  ---cett-----at---agacagctcgaaacccccatagacett--tato
VIRAL ENVELOPE PROTEIN 46      L      I      D S S N P H R P      L S
                        |  ||      ||  || || | || | || | || |
Sbjct:          5743  agacgttctgtctgggacacaaaggcag-tccagcccccttggga-cttggtggc
env protein      65      D V V W D T K A V Q P P W T W W

```

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi?0>

22/02/01

## Page 2 of 5

```

Query:                6230 atgagctggggaat-agttttttattatatg-g-cggggga-gcagg-gtcc
VIRAL ENVELOPE PROTEIN 204 Y E L G N S F L L Y G G G A G S
                        | | | | | | | | | | | | | | | | | | | |
Sbjct:                6254 aaaacctggggattaag-----attctatgtgtctggacatccaggcgta
env protein           235 K T W G L R F Y V S G H P G V

```

22/02/01